What Is Claimed:

1. A method comprising:

receiving a first uniform resource locator (URL) including one or more parameters;

retrieving content corresponding to the first URL;

retrieving content corresponding to a plurality of URLs having different parameter combinations of the one or more parameters;

identifying a parameter combination from the plurality of URLs that corresponds to content that is approximately the same as the content corresponding to the first URL; and

generating one or more URL rewrite rules based on the identified parameter combination.

2. The method of claim 1, wherein the different parameter combinations include the first URL with no parameters, the first URL with each of the one or more parameters individually, and the first URL with combinations of the one or more parameters.

3. The method of claim 1, further comprising:

performing the receiving a first URL, retrieving content corresponding to the first URL, retrieving content corresponding to the plurality of URLs, and identifying the parameter combination, for multiple different first URLs that each include the same parameters; and

generating the one or more URL rewrite rules for the identified parameter combinations.

4. The method of claim 3, wherein the rewrite rules specify that parameters that do not occur in a threshold number of the identified parameter combinations are to be removed.

5. The method of claim 1, wherein each rewrite rule applies to a particular web site or web host.

6. The method of claim 1, wherein the reduced number of parameters includes a minimum number of parameters.

7. A method for converting a uniform resource locator (URL) into a canonical form of the URL, the method comprising:

receiving a URL that refers to content and that contains a parameter set including at least one parameter;

applying a predetermined rewrite rule to the URL that removes the at least one parameter from the URL when the at least one parameter does not affect the content referred to by the URL; and

outputting the rewritten URL as the canonical form of the URL.

8. The method of claim 7, wherein the predetermined rewrite rule is determined by:

receiving a plurality of URLs that contain the parameter set; and

identifying parameters in the parameter set that do not contribute to content.

9. The method of claim 8, wherein the identifying parameters in the parameter set that do contribute to content includes retrieving content corresponding to a sampled URL containing combinations of parameters in the parameter set and identifying a combination of parameters for which the retrieved content is approximately the same as the content corresponding to the parameter set and that contains a reduced number of parameters.

10. The method of claim 9, wherein the combinations of parameters include the sampled URL with no parameters, the sampled URL with individual parameters, and the sampled URL with combinations of the at least one parameter.

11. The method of claim 7, wherein the rewrite rule applies to a particular web site or web host.

12. One or more devices comprising:

at least one fetch bot configured to download content on a network from locations specified by uniform resource locators (URLs);

a content manager configured to extract URLs from the downloaded content;

a rewrite component configured to

receive a URL that refers to content and that contains a parameter set including at least one parameter,

apply a predetermined rewrite rule to the URL that removes the at least one parameter from the URL when the at least one parameter does not affect the content referred to by the URL,

output the rewritten URL as the canonical form of the URL; and

a URL manager configured to store the canonical form of the URL.

13. The one or more devices of claim 12, wherein the predetermined rewrite rule is determined by:

receiving a plurality of URLs that contain the parameter set; and

identifying parameters in the parameter set that do not contribute to content.

14. The one or more devices of claim 13, wherein the identifying parameters in the parameter set that do contribute to content includes retrieving content corresponding to a sampled URL containing combinations of parameters in the parameter set and identifying a combination of parameters for which the

retrieved content is approximately the same as the content corresponding to the parameter set and that contains a minimum number of parameters.

15. The one or more devices of claim 14, wherein the combinations of parameters include the sampled URL with no parameters, the sampled URL with individual parameters, and the sampled URL with combinations of the at least one parameter.

16. The one or more devices of claim 12, wherein each rewrite rule applies to a particular web site or web host.

17. A system comprising:

means for receiving a first uniform resource locator (URL) including one or more parameters;

means for retrieving content corresponding to the first URL;

means for retrieving content corresponding to a plurality of URLs having different parameter combinations of the one or more parameters;

means for identifying the parameter combination from the plurality of URLs that corresponds to content that is approximately the same as the content corresponding to the first URL and that contains a minimum number of parameters; and

means for generating one or more URL rewrite rules based on the identified parameter combination.

18. A computer-readable medium including programming instructions executed by a processor, the programming instructions comprising:

instructions for receiving a first uniform resource locator (URL) including one or more parameters;

instructions for retrieving content corresponding to the first URL;

instructions for retrieving content corresponding to a plurality of URLs having different parameter combinations of the one or more parameters;

instructions for identifying the parameter combination from the plurality of URLs that corresponds to content that is approximately the same as the content corresponding to the first URL and that contains a minimum number of parameters; and

instructions for generating one or more URL rewrite rules based on the identified parameter combination.